ON THE CHARACTERISTIC FUNCTIONS OF THE DISTRIBUTIONS OF ESTIMATES OF VARIOUS DEVIATIONS IN SAMPLES FROM A NORMAL POPULATION

By M. Kac

Cornell University

1. Summary. An explicit formula for the characteristic function of the deviation

$$\frac{1}{n}\sum_{k=1}^{n} | X_k - \bar{X} |^{\alpha}, \qquad \alpha > 0,$$

is derived for samples from a normal population. For $\alpha = 1$ one can calculate the probability density function but the result does not seem to be in complete agreement with a recent formula of Goodwin [1].

2. Introduction. Let $X_1, X_2, \cdots, X_n$ be independent, normally distributed random variables each having mean 0 and variance 1.

Let, as usual,

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n},$$

and denote by $Y_n(\alpha)$ the deviation

(1) $$Y_n(\alpha) = \frac{1}{n}\sum_{k=1}^{n} | X_k - \bar{X} |^{\alpha}, \qquad \alpha > 0.$$

The purpose of this note is to show that

(2) $$F_n(\xi) = E\{\exp (i\xi Y_n(\alpha))\}$$
$$= \frac{1}{\sqrt{n}(\sqrt{2\pi})^{n+1}} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} e^{-x^2/2} e^{i/n(\xi|x|^{\alpha}+\eta x)}\, dx \right]^n d\eta.$$

It is easy to check that for $\alpha = 2$ one obtains the well known expression

$$\left(1 - \frac{2i\xi}{n}\right)^{-(n-1)/2}$$

Moreover, if $\alpha = 1$ one can actually find the probability density of $Y_n(1)$. The resulting expression is fairly complicated and it strongly resembles an expression recently obtained by Goodwin [1]. Except for the relatively simple case $n = 3$, it does not seem easy to verify that our formula is equivalent to that of Goodwin.

Although deviations corresponding to values of α different from 1 and 2 are of little practical value the explicit formula (2) may be of some interest. It is also hoped that the method of derivation may prove useful in other cases.

3. The derivation of (2). We start with the observation that

$$\bar{X} \text{ and } Y_n(\alpha)$$

are statistically independent (see e.g. Daly [2]).

Denote by

$$E^*\{\, |\bar{X}| < \epsilon, \exp(i\xi Y_n(\alpha))\}$$

the integral of $\exp(i\xi Y_n(\alpha))$ extended over that portion of the sample space in which $|\bar{X}| < \epsilon$. Thus the conditional expectation $E\{\exp(i\xi Y_n(\alpha)) \mid |\bar{X}| < \epsilon\}$ is given by the formula

$$E\{\exp(i\xi X_n(\alpha)) \mid |\bar{X}| < \epsilon\} = \frac{E^*\{\, |\bar{X}| < \epsilon, \exp(i\xi Y_n(\alpha))\}}{\text{Prob}\{\, |\bar{X}| < \epsilon\}}.$$

Because of the independence of $\bar{X}$ and $Y_n(\alpha)$ we have

$$E\{\exp(i\xi Y_n(\alpha))\} = \frac{E^*\{\, |\bar{X}| < \epsilon, \exp(i\xi Y_n(\alpha))\}}{\text{Prob}\{\, |\bar{X}| < \epsilon\}}. \tag{3}$$

For the sake of simplicity we assume now that $\alpha \geq 1$ and note that

$$\left| \exp(i\xi Y_n(\alpha)) - \exp\left(\frac{i\xi}{n}\sum_1^n |X_k|^\alpha\right)\right| \leq \frac{\xi}{n}\sum_1^n \big|\, |X_k|^\alpha - |X_k - \bar{X}|^\alpha \big|$$

$$\leq \frac{\alpha\xi |\bar{X}|}{n}\sum_1^n (|X_k| + |\bar{X}|)^{\alpha-1}.$$

Thus, on the portion of the sample space where $|\bar{X}| < \epsilon$, we have

$$\left| \exp(i\xi Y_n(\alpha)) - \exp\left(\frac{i\xi}{n}\sum_1^n |X_k|^\alpha\right)\right| \leq \frac{\alpha\xi\epsilon}{n}\sum_1^n (|X_k| + \epsilon)^{\alpha-1}$$

and consequently

$$\left| E^*\{\, |\bar{X}| < \epsilon, \exp(i\xi Y_n(\alpha))\} - E^*\left\{ |\bar{X}| < \epsilon, \exp\left(\frac{i\xi}{n}\sum_1^n |X_k|^\alpha\right)\right\}\right|$$

$$\leq \frac{\alpha\xi\epsilon}{n} E^*\left\{ |\bar{X}| < \epsilon, \sum_1^n (|X_k| + \epsilon)^{\alpha-1}\right\}.$$

Clearly $E^*\left\{ |\bar{X}| < \epsilon, \sum_1^n (|X_k| + \epsilon)^{\alpha-1}\right\}$, approaches 0, as ϵ approaches 0, hence by (3)

$$E\{\exp(i\xi Y_n(\alpha))\} = \lim_{\epsilon \to 0} \frac{E^*\left\{ |\bar{X}| < \epsilon, \exp\left(\frac{i\xi}{n}\sum_1^n |X_k|^\alpha\right)\right\}}{\text{Prob}\{\, |\bar{X}| < \epsilon\}}. \tag{4}$$

Using the fact that

$$\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\sin \epsilon\eta}{\eta}\exp(i\eta\bar{X})\,d\eta \begin{cases} = 1, & |\bar{X}| < \epsilon, \\ = \tfrac{1}{2}, & |\bar{X}| = \epsilon, \\ = 0, & |\bar{X}| > \epsilon, \end{cases}$$

we obtain easily

$$(5)\qquad \begin{aligned} &E^*\left\{ |\bar{X}| < \epsilon, \exp\left(\frac{i\xi}{n}\sum_1^n |X_k|^\alpha\right)\right\} \\ &\qquad = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\sin\epsilon\eta}{\eta} E\left\{\exp\frac{i}{n}\sum_1^n (\xi\,|X_k|^\alpha + \eta X_k)\right\} d\eta. \end{aligned}$$

The justification of interchanging of the order of integration (from $-\infty$ to ∞) and the operation E can be made quite simply (see e.g. Kac and Steinhaus [3]). Notice now that

$$\begin{aligned} &E\left\{\exp\frac{i}{n}\sum_1^n (\xi\,|X_k|^\alpha + \eta\, X_k)\right\} \\ &\qquad = \left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(-\frac{x^2}{2}\right)\exp\frac{i}{n}(\xi\,|x|^\alpha + \eta x)\,dx\right]^n = \varphi_n(\xi, \eta) \end{aligned}$$

and that $\varphi_n(\xi, \eta)$ is absolutely integrable in $(-\infty, \infty)$ as a function of η. Thus, as $\epsilon \to 0$,

$$(6)\qquad \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\sin\epsilon\eta}{\eta}\varphi_n(\xi, \eta)\,d\eta \sim \frac{\epsilon}{\pi}\int_{-\infty}^{\infty}\varphi_n(\xi, \eta)\,d\eta.$$

Furthermore (as $\epsilon \to 0$)

$$(7)\qquad \text{Prob}\{ |\bar{X}| < \epsilon \} \sim 2\epsilon\frac{\sqrt{n}}{\sqrt{2\pi}}$$

and combining this with (6), (5) and (4) we get

$$(8)\qquad E\{\exp(i\xi Y_n(\alpha))\} = \frac{1}{\sqrt{2\pi n}}\int_{-\infty}^{\infty}\varphi_n(\xi, \eta)\,d\eta.$$

This, of course, is equivalent to (2).

4. Density function of the mean deviation. If $\alpha = 1$ one can obtain an expression for the probability density $f_n(\beta)$ of $Y_n(\alpha)$. We note first that

$$\begin{aligned} &\int_{-\infty}^{\infty}\exp\left(-\frac{x^2}{2}\right)\exp\frac{i}{n}(\xi\,|x| + \eta x)\,dx \\ &\qquad = n\int_0^{\infty}\exp\left(-\frac{n^2x^2}{2}\right)\exp i(\xi + \eta)x\,dx \\ &\qquad + n\int_0^{\infty}\exp\left(-\frac{n^2x^2}{2}\right)\exp i(\xi - \eta)x\,dx = n\{R(\xi + \eta) + R(\xi - \eta)\} \end{aligned}$$

where

$$R(u) = \int_0^\infty \exp\left(-\frac{n^2 x^2}{2}\right) \exp(iux)\, dx.$$

Using (2) (with $\alpha = 1$) we obtain

$$F_n(\xi) = \frac{n^n}{\sqrt{n}(\sqrt{2\pi})^{n+1}} \int_{-\infty}^{\infty} \left[\sum_{k=0}^{n} \binom{n}{k} R^k(\xi + \eta) R^{n-k}(\xi - \eta)\right] d\eta.$$

Let us first look at the summands corresponding to $k = 0$ and $k = \text{n}$. We have

$$\int_{-\infty}^{\infty} R^n(\xi - \eta)\, d\eta = \int_{-\infty}^{\infty} R^n(\eta)\, d\eta = \int_{-\infty}^{\infty} R^n(\xi + \eta)\, d\eta.$$

Now, $R(\eta)$ is the Fourier transform of

$$\zeta(x) = \begin{cases} 0, & x < 0, \\ \exp\left(-\dfrac{n^2 x^2}{2}\right), & x > 0, \end{cases}$$

and hence $R^n(\eta)$ is the Fourier transform of the convolution

$$\underbrace{\zeta * \zeta * \cdots * \zeta}_{n} = \zeta^{(n)}(x).$$

It is easily seen (using integration by parts) that

$$R(\eta) = O\left(\frac{1}{|\eta|}\right)$$

for large $|\eta|$ and hence for $n \geq 2$, $R^n(\eta)$ is absolutely integrable in $(-\infty, \infty)$. It follows (classical inversion formula) that

$$\int_{-\infty}^{\infty} R^n(\eta)\, d\eta = 2\pi\zeta^{(n)}(0).$$

Since for $n \geq 2$, $\zeta^{(n)}(x)$ is continuous and $\zeta^{(n)}(x) = 0$ for $x < 0$ we have $\zeta^{(n)}(0) = 0$. Thus

$$F_n(\xi) = \frac{n^{n-\frac{1}{2}}}{(\sqrt{2\pi})^{n+1}} \sum_{k=1}^{n-1} \binom{n}{k} \int_{-\infty}^{\infty} R^k(\xi + \eta) R^{n-k}(\xi - \eta)\, d\eta.$$

It is fairly easy to check that

$$\int_{-\infty}^{\infty} R^k(\xi + \eta) R^{n-k}(\xi - \eta)\, d\eta = \pi \int_{-\infty}^{\infty} \exp(i\xi x) \zeta^{(k)}\left(\frac{x}{2}\right) \zeta^{(n-k)}\left(\frac{x}{2}\right) dx$$

so that

$$F_n(\xi) = \frac{\pi n^{n-\frac{1}{2}}}{(\sqrt{2\pi})^{n+1}} \int_{-\infty}^{\infty} \exp(i\xi x) \sum_{k=1}^{n-1} \binom{n}{k} \zeta^{(k)}\left(\frac{x}{2}\right) \zeta^{(n-k)}\left(\frac{x}{2}\right) dx.$$

Finally,

$$f_n(\beta) = \frac{\pi n^{n-\frac{1}{2}}}{(\sqrt{2\pi})^{n+1}} \sum_{k=1}^{n-1} \binom{n}{k} \zeta^{(k)}\left(\frac{\beta}{2}\right) \zeta^{(n-k)}\left(\frac{\beta}{2}\right).$$

I have not been able, except for $n = 3$, to verify directly that this formula is identical with that of Goodwin.

REFERENCES

[1] H. J. GOODWIN, "On the distribution of the estimate of mean deviation obtained from samples from a normal population," *Biometrika*, Vol. 33 (1945), pp. 254–256.
[2] J. F. DALY, "On the use of the sample range in an analogue of Student's *t*-test," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 71–74.
[3] M. KAC AND H. STEINHAUS, "Sur les fonctions indépendantes III," *Stud. Math.*, Vol. 6 (1936), pp. 93–94.